In this lecture, we will discuss the basics of quantum information theory. In particular, we will discuss mixed quantum states, density matrices, von Neumann entropy and the trace distance between mixed quantum states.

1 Mixed Quantum State

So far we have dealt with *pure* quantum states

$$|\psi\rangle = \sum_x \alpha_x |x\rangle.$$

This is not the most general state we can think of. We can consider a probability distribution of pure states, such as $|0\rangle$ with probability $1/2$ and $|1\rangle$ with probability $1/2$. Another possibility is the state

$$\begin{cases} |+\rangle = \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle) & \text{with probability } 1/2 \\ |-\rangle = \frac{1}{\sqrt{2}}(|0\rangle - |1\rangle) & \text{with probability } 1/2 \end{cases}$$

In general, we can think of *mixed* state as a collection of pure states $|\psi_i\rangle$, each with associated probability p_i, with the conditions $0 \le p_i \le 1$ and $\sum_i p_i = 1$. One reason we consider such mixed states is because the quantum states are hard to isolate, and hence often entangled to the environment.

2 Density Matrix

Now we consider the result of measuring a mixed quantum state. Suppose we have a mixture of quantum states $|\psi_i\rangle$ with probability p_i. Each $|\psi_i\rangle$ can be represented by a vector in $\mathscr{C}^{2^n}$, and thus we can associate the outer product $|\psi_i\rangle\langle\psi_i| = \psi_i \psi_i^*$, which is an $2^n \times 2^n$ matrix

$$\begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_N \end{pmatrix} \begin{pmatrix} \bar{a}_1 & \bar{a}_2 & \cdots & \bar{a}_N \end{pmatrix} = \begin{pmatrix} a_1\bar{a}_1 & a_1\bar{a}_2 & \cdots & a_1\bar{a}_N \\ a_2\bar{a}_2 & a_1\bar{a}_2 & \cdots & a_2\bar{a}_N \\ \vdots & & & \vdots \\ a_N\bar{a}_1 & a_N\bar{a}_2 & \cdots & a_N\bar{a}_N \end{pmatrix}.$$

We can now take the average of these matrices, and obtain the *density matrix* of the mixture $\{p_i, |\psi_i\rangle\}$:

$$\rho = \sum_i p_i |\psi_i\rangle\langle\psi_i|.$$

We give some examples. Consider the mixed state $|0\rangle$ with probability of $1/2$ and $|1\rangle$ with probablity $1/2$. Then

$$|0\rangle\langle 0| = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \begin{pmatrix} 1 & 0 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix},$$

and

$$|1\rangle\langle 1| = \begin{pmatrix} 0 \\ 1 \end{pmatrix} \begin{pmatrix} 0 & 1 \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}.$$

Thus in this case

$$\rho = \frac{1}{2}|0\rangle\langle 0| + \frac{1}{2}|1\rangle\langle 1| = \begin{pmatrix} 1/2 & 0 \\ 0 & 1/2 \end{pmatrix}.$$

Now consider another mixed state, this time consisting of $|+\rangle$ with probability $1/2$ and $|-\rangle$ with probability $1/2$. This time we have

$$|+\rangle\langle+| = (1/2)\begin{pmatrix} 1 \\ 1 \end{pmatrix}\begin{pmatrix} 1 & 1 \end{pmatrix} = \frac{1}{2}\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix},$$

and

$$|-\rangle\langle-| = (1/2)\begin{pmatrix} 1 \\ -1 \end{pmatrix}\begin{pmatrix} 1 & -1 \end{pmatrix} = \frac{1}{2}\begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix}.$$

Thus in this case the offdiagonals cancel, and we get

$$\rho = \frac{1}{2}|+\rangle\langle+| + \frac{1}{2}|-\rangle\langle-| = \begin{pmatrix} 1/2 & 0 \\ 0 & 1/2 \end{pmatrix}.$$

Note that the two density matrices we computed are identical, even though the mixed state we started out was different. Hence we see that it is possible for two different mixed states to have the same density matrix.

We now show that two mixed states can be distinguished if and only if the density matrix ρ are different:

Theorem 16.1: *Suppose we measure a mixed state $\{p_j, |\psi_j\rangle\}$ in an orthonormal bases $|\beta_k\rangle$. Then the outcome is $|\beta_k\rangle$ with probability $\langle\beta_k|\rho|\beta_k\rangle$.*

Proof: We denote the probability of measuring $|\beta_k\rangle$ by $\Pr[k]$. Then

$$\begin{aligned} \Pr[k] &= \sum_j p_j |\langle\psi_j|\beta_k\rangle|^2 \\ &= \sum_j p_j \langle\beta_k|\psi_j\rangle\langle\psi_j|\beta_k\rangle \\ &= \left\langle \beta_k \middle| \sum_j p_j |\psi_j\rangle\langle\psi_j| \middle| \beta_k \right\rangle \\ &= \langle\beta_k|\rho|\beta_k\rangle. \end{aligned}$$

$\square$

corollary If we measure the mixed state $\{p_j, |\psi_j\rangle\}$ in the standard basis, we have $\Pr[k] = \rho_{k,k}$, the diagonal entry of the density matrix ρ.

We list several more properties of the density matrix:

1. $\mathrm{tr}\rho = 1$. This follows immediately from Corollary 2, since the probabilities $\Pr[k]$ must add up to 1.
2. ρ is Hermitian. This follows from the fact that ρ is a sum of Hermitian outer products ($(\psi\psi^*)^* = \psi\psi^*$).
3. Eigenvalues of ρ are non-negative. First of all, eigenvalues of a Hermitian matrix is real. Suppose that λ and $|e\rangle$ are corresponding eigenvalue and eigenvector. Then if we measure in the eigenbasis, we have

   $$\Pr[e] = \langle e|\rho|e\rangle = \lambda\langle e|e\rangle = \lambda.$$

   Since the probability must be non-negative, we see that $\lambda \geq 0$.

Now suppose we have two mixed states, with density matrices A and B such that $A \neq B$. We can ask, what is a good measurement to distinguish the two states? We can diagonalize the difference $A-B$ to get $A-B = E\Lambda E^*$, where E is the matrix of orthogonal eigenvectors. Then if e_i is an eigenvector with eigenvalue λ_i, then λ_i is the difference in the probability of measuring e_i:

$$\Pr_A[i] - \Pr_B[i] = \lambda_i.$$

We can define the distance between two probability distributions (with respect to a basis E) as

$$\left|\mathscr{D}_A - \mathscr{D}_B\right|_E = \sum \left(\Pr_A[i] - \Pr_B[i]\right).$$

If E is the eigenbasis, then

$$\left|\mathscr{D}_A - \mathscr{D}_B\right|_E = \sum_i |\lambda_i| = \mathrm{tr}|A-B| = \|A-B\|_{\mathrm{tr}},$$

which is called the trace distance between A and B.

claim Measuring with respect to the eigenbasis E (of the matrix $A-B$) is optimal in the sense that it maximizes the distance $|\mathscr{D}_A - \mathscr{D}_B|_E$ between the two probability distributions.

Before we prove this claim, we introduce the following definition and lemma without proof. **definition** Let $\{a_i\}_{i=1}^N$ and $\{b_i\}_{i=1}^N$ be two non-increasing sequences such that $\sum_i a_i = \sum_i b_i$. Then the sequence $\{a_i\}$ is said to majorize $\{b_i\}$ if for all k,

$$\sum_{i=1}^k a_i \geq \sum_{i=1}^k b_i.$$

lemma[Schur] Eigenvalues of any Hermitian matrix majorizes the diagonal entries (if both are sorted in nonincreasing order).

Now we can prove claim 2.

proof Since we can reorder the eigenvectors, we can assume $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$. Note that $\mathrm{tr}(A-B) = 0$, so we must have $\sum_i \lambda_i = 0$. We can split the λ_i's into two groups: positive ones and negative ones, we must have

$$\sum_{\lambda_i > 0} = \frac{1}{2}\|A-B\|_{\mathrm{tr}} \qquad \sum_{\lambda_i < 0} = -\frac{1}{2}\|A-B\|_{\mathrm{tr}}.$$

Thus

$$\max_k \sum_{i=1}^k \lambda_i = \frac{1}{2}\|A-B\|_{\mathrm{tr}}.$$

Now consider measuring in another basis. Then the matrix $A-B$ is represented as $H = F(A-B)F^*$, and let $\mu_1 \geq \mu_2 \geq \cdots \geq \mu_n$ be the diagonal entries of H. Similar argument shows that

$$\max_k \sum_{i=1}^k \mu_i = \frac{1}{2}\sum_{i=1}^n |\mu_i| = \frac{|\mathscr{D}_A - \mathscr{D}_B|_F}{2}.$$

But by Schur's lemma the λ_i's majorizes μ_i's, so we must have

$$|\mathscr{D}_A - \mathscr{D}_B|_F \leq |\mathscr{D}_A - \mathscr{D}_B|_E = \|A-B\|_{\mathrm{tr}}.$$

3 Von Neumann Entropy

Consider the following two mixtures and their density matrices:

$$\left.\begin{array}{lll} \cos\theta|0\rangle+\sin\theta|1\rangle \quad \text{w.p. } 1/2 & =\frac{1}{2}\begin{pmatrix} c\theta \\ s\theta \end{pmatrix}\begin{pmatrix} c\theta & s\theta \end{pmatrix} & =\frac{1}{2}\begin{pmatrix} c^2\theta & c\theta s\theta \\ c\theta s\theta & s^2\theta \end{pmatrix} \\ \cos\theta|0\rangle-\sin\theta|1\rangle \quad \text{w.p. } 1/2 & =\frac{1}{2}\begin{pmatrix} c\theta \\ -s\theta \end{pmatrix}\begin{pmatrix} c\theta & -s\theta \end{pmatrix} & =\frac{1}{2}\begin{pmatrix} c^2\theta & -c\theta s\theta \\ -c\theta s\theta & s^2\theta \end{pmatrix} \end{array}\right\} = \begin{pmatrix} \cos^2\theta & 0 \\ 0 & \sin^2\theta \end{pmatrix}$$

$$\left.\begin{array}{lll} |0\rangle \quad \text{w.p.}\cos^2\theta & =\cos^2\begin{pmatrix} 1 \\ 0 \end{pmatrix}\begin{pmatrix} 1 & 0 \end{pmatrix} & =\cos^2\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \\ |1\rangle \quad \text{w.p. } \sin^2\theta & =\sin^2\theta\begin{pmatrix} 0 \\ 1 \end{pmatrix}\begin{pmatrix} 0 & 1 \end{pmatrix} & =\sin^2\begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} \end{array}\right\} = \begin{pmatrix} \cos^2\theta & 0 \\ 0 & \sin^2\theta \end{pmatrix}$$

Thus, since the mixtures have identical density matrices, they are indistinguishable.

Let $H(X)$ be the *Shannon Entropy* of a random variable X which can take on states $p_1 \dots p_n$.

$$H(\{p_i\}) = -\sum_i \log\frac{1}{p_i} = 1$$

In the quantum world, we define an analogous quantity, $S(\rho)$, the *Von Neumann entropy* of a quantum ensemble with density matrix ρ.

$$S(\rho) = H\{\cos^2\theta, \sin^2\theta\} = \cos^2\theta\rho\frac{1}{\cos^2\theta} + \sin^2\theta\rho\frac{1}{\sin^2\theta}$$

For the set of quantum states $\psi_1 \dots \psi_n$ with probabilities $p_1 \dots_n$ and density matrix ρ, we can diagonalize ρ, letting π_i be the i^{th} element along the diagonal. This yields

$$p_i, |\psi_i\rangle \equiv \lambda_i |e_i\rangle$$

We can also express the Von Neumann entropy in terms of the diagonal elements, as

$$S(\rho) = H(\{\pi_i\})$$